Census and Sample transcript

Animation: A train travels across the screen. The background is a cityscape at night.

Script: There are 100 people on this train.

Animation: 100 white silhouettes of people appear. 50 are male and 50 are female. They are arranged into 5 rows of 20 people each.

Script: The population of interest is all people in the train.

We want to find out what people do when they are on the train.

Let's take a census of the train, and collect data on what people were doing while on the train.

Animation: The 100 white silhouettes of people reappear.

Script: 50 were listening to music (50%)

Animation: 50 of the silhouettes turn yellow

Script: 25 were playing on smart phones (25%)

Animation: 25 of the silhouettes turn orange

Script: 15 were reading (15%)

Animation: 15 of the silhouettes turn green

Script: 10 were sleeping (10%)

Animation: The remaining 10 silhouettes turn blue

Script: We have taken a complete enumeration because we have collected data from every unit in the population.

Animation: A column graph appears presenting the results as yellow columns. It is titled "Main activity of train passengers". The graph represents a "Census (population 100, count = 100)".

Script: Now let's take a sample of the population of interest. A sample is a subset of the population which is selected to represent the whole population.

Let's take a sample of 50 people.

Animation: 100 white silhouettes are shown again. 50 of them fade away

Script: 26 were listening to music (52%)

Animation: 26 of the 50 bright silhouettes turn yellow

Script: 13 were playing on smart phones (26%)

Animation: 13 of the 50 bright silhouettes turn orange

Script: 8 were reading (16%)

Animation: 8 of the 50 bright silhouettes turn green

Script: 3 were sleeping (6%)

Animation: 3 of the 50 bright silhouettes turn blue.

A column graph is shown presenting the results of the sample as orange columns. It is titled "Main activity of train passengers". The graph represents the sample (population 100, count = 50).

Script: The results of the sample were quite representative of the population of the train.

Animation: The two column graphs for census and sample are combined, with the yellow census columns presented alongside the orange sample columns. The orange and yellow columns are similar in height.

Script: We can see from the graph that the sample we drew represented our population quite well.

However, it is also possible to draw different samples from the same population and get quite different results.

For example, we could take a sample focussed on a group of school children at one end of the train.

Sample 2:

Listening to music = 35 (70%)

Playing on smart phones = 14 (28%)

Reading = 1 (2%)

Sleeping = 0 (0%)

Animation: A column graph appears presenting the results as red columns. It is titled "Main activity of train passengers". The red is shown as representing "Sample 2" (population 100, count = 50).

A new graph appears which combines the three datasets, with the yellow census columns presented alongside the orange sample 1 and the red sample 2 columns. The orange and yellow columns are similar in height, but the red column is much greater for listening to music and much less for reading and sleeping.

Script: In this instance, the sample is less representative of the whole population of the train.

The graph shows that the sample over-represents passengers listening to music or playing on smart phones and under-represents those reading or sleeping.

As we saw, samples can represent the population well but are also subject to error and variability.

This should be considered when using data from a sample survey.

Animation: End